ELSEVIER

Ecological Indicators 6 (2006) 589–608

ECOLOGICAL INDICATORS

This article is also available online at:
www.elsevier.com/locate/ecolind

Neural net modeling of estuarine indicators: Hindcasting phytoplankton biomass and net ecosystem production in the Neuse (North Carolina) and Trout (Florida) Rivers, USA

David F. Millie [a,f,*], Gary R. Weckman [b], Hans W. Paerl [c], James L. Pinckney [d], Brian J. Bendis [e], Ryan J. Pigg [f], Gary L. Fahnenstiel [g]

[a] *Florida Institute of Oceanography, University of South Florida, Bayboro Campus, c/o Fish & Wildlife Research Institute, Florida Fish & Wildlife Conservation Commission, 100 Eighth Avenue, S.E., Saint Petersburg, FL 33701, USA*
[b] *Department of Industrial & Manufacturing Systems Engineering, Ohio University, Athens, OH 45701, USA*
[c] *Institute of Marine Sciences, University of North Carolina, Chapel Hill, Morehead City, NC 28557, USA*
[d] *Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA*
[e] *AMJ Equipment Corporation, Lakeland, FL 33815, USA*
[f] *Fish & Wildlife Research Institute, Florida Fish & Wildlife Conservation Commission, Saint Petersburg, FL 33701, USA*
[g] *National Oceanic & Atmospheric Administration, Great Lakes Environmental Research Laboratory, Lake Michigan Field Station, Muskegon, MI 49441, USA*

Received 3 June 2005; received in revised form 15 August 2005; accepted 16 August 2005

Abstract

Phytoplankton biomass, as chlorophyll (Chl) *a*, and net ecosystem production (NEP), were modeled using artificial neural networks (ANNs). Chl *a* varied seasonally and along a saline gradient throughout the Neuse River (North Carolina). NEP was extremely dynamic in the Trout River (Florida), with phototrophic or heterotrophic conditions occurring over short-term intervals. Physical and chemical variables, arising from meteorological and hydrological conditions, created spatial and/or temporal gradients in both systems and served as interacting predictors for the trends/patterns of Chl *a* and NEP. ANNs outperformed comparable linear regression models and reliably modeled Chl *a* concentrations less than 20 $\mu g\ L^{-1}$ and NEP values, denoting the apparent non-linear interactions among abiotic and indicator variables. ANNs underestimated Chl *a* concentrations greater than 20 $\mu g\ L^{-1}$, likely due to the periodicity of data acquisition not being sufficient to generalize system variability, the designated 'lag' effect for variables not being adequate to portray estuarine flow dynamics, the exclusion of (one or more) variables that would have improved prediction, and/or an unrealistic expectation of network performance. Variables indicative of meteorological and hydrological forcing and/or proxy measurements of phytoplankton had the greatest relative impact on prediction of Chl *a* and NEP. Except for their predictive capability, ANNs might appear to be of limited value for ecological applications and problem solving; interpreting the absolute impact of and/or interacting relationships among network

* Corresponding author. Tel.: +1 727 896 8626; fax: +1 727 550 4222.
E-mail address: david.millie@myfwc.com (D.F. Millie).

1470-160X/$ – see front matter
doi:10.1016/j.ecolind.2005.08.021

variables is intrinsically difficult. Statistical methods or 'rule extraction' algorithms that convey comprehensible network interpretation are needed prior to the routine use of ANNs in programs assessing and/or forecasting the response of biotic indicators to perturbation or for a means to discern estuarine function.

Keywords: Algae; Estuary; Neural networks; Regression; Ecosystem modeling

1. Introduction

Characterization based on ecological indicators is central to assessing the status or 'health' of coastal systems in response to anthropogenic stressors and/or natural disturbances. The chlorophyll (Chl) *a* concentration of a water column is a universally accepted indicator for total phytoplankton biomass and a useful estimate for a population's response to changing environmental/endogenous variables and/or system-level eutrophication (e.g. Millie et al., 1993; Harris, 1994, 1996; Paerl et al., 2003; Soyupak and Chen, 2004). Net ecosystem production (NEP) represents the balance between production and respiration and as such, is a proxy for system trophic state; positive values indicate that autochthonous production of organic matter dominates (phototrophy) whereas negative values signify that allochthonous sources are most influential (heterotrophy; see Odum, 1956; Swaney et al., 1999; Caffrey et al., 1998; Smith and Hollibaugh, 1993; Caffrey, 2003, 2004).

It is imperative that we not only understand how coastal systems function, but also predict how they will be affected by change (National Biological Information Infrastructure; http://www.nbii.gov/index.html). However, our ability to predict population- and/or system-level response to stressors and/or disturbances within estuarine waters is poor, due in part to the inability to both functionally link and conceptually model the interactions between abiotic and biotic variables (Paerl, 1988; Rudek et al., 1991; Glibert et al., 1995; Sigua et al., 2000; Cloern, 2001; Paerl et al., 2005). Clearly, accurate assessment of chronic and episodic perturbation requires the identification, quantification, and interpretation of integrative indicators capable of coupling population/community structure to ecosystem integrity (Committee on Environmental and Natural Resources, 1997; Bortone, 2005; Jordon and Smith, 2005; Paerl et al., 2005).

The ''scaling up'' of indicator data for forecasting the consequences (and controls) of estuarine perturbation requires diverse and robust modeling approaches (after Barnes and Mazzotti, 2005; Marshall, 2005; Paerl et al., 2005). Although deterministic models (based on physical/chemical/biological relationships) have a wide range of applicability and can cope with deviations within the system modeled, they often are complex, require extensive data sets, and contain numerous parameters whose values are uncertain and/or require initialization (Maier et al., 1998; Murray and Parslow, 1999; Walsh et al., 2001). Linear and non-linear statistical models approximate data relationships solely through mathematical functions and as such, require no theoretical 'biological guidelines'. Due to the non-linear and often stochastic interactions that commonly exist among patterns and processes in aquatic systems (see Smith et al., 1988; Harris, 1994; Mazumder, 1994), artificial neural networks (ANNs) have become increasingly popular in modeling phytoplankton abundance and production (e.g. Recknagel et al., 1997; Barciela et al., 1999; Maier et al., 1998; Scardi and Harding, 1999; Olden, 2000; Richardson et al., 2002; Gurbuz et al., 2003; Lee et al., 2003). ANNs are non-linear parametric models that reproduce correlated patterns between/among variables through repetitive data processing. In contrast to linear models, they do not require a known probability distribution of variables and easily accommodate large 'noisy' data sets reflecting seasonal and cyclic variation (Smith and Mason, 1997; Maier et al., 1998; Richardson et al., 2002).

Although ANNs show promise for modeling ecological indicators within dynamic coastal waters, their application is relatively new and requires validation and interpretation across diverse systems. Here, ANNs modeled Chl *a* and NEP in distinct

estuaries within the southeastern USA. Specifically, we: (1) identified spatial/temporal patterns of abiotic predictor variables, Chl *a*, and NEP; (2) developed and validated networks for realistic prediction of Chl *a* concentrations, NEP values, and phototrophic/heterotrophic classifications; and (3) elucidated the (relative) importance of predictor variables within the networks. The potential usefulness and shortcomings of using ANNs for modeling estuarine indicators within coastal assessment programs also are discussed.

2. Methods

2.1. Study sites and data acquisition

The Neuse River flows through ca. 300 km of North Carolina's most productive and rapidly expanding urban, industrial and agricultural regions before emptying into Pamlico Sound, the lower portion of the Albermarle–Pamlico Estuarine System (Fig. 1). Changing land-use activities in the watershed coupled with major climatic perturbations have generated various non-point source nutrient and sediment inputs into the estuarine portion of the Neuse River, ultimately promoting episodic proliferation of nuisance/harmful algal blooms, hypoxia/anoxia events, and alterations in biogeochemical cycling and microbial, invertebrate, and fish community structure and function (Pinckney et al., 1997, 1999; Paerl et al., 2001, 2003, 2005; Cooper et al., 2004). From April 1994 through December 2003, water-quality data (Table 1) were acquired bi- to tri-weekly along the length of the estuary at (up to) 12 sites, ranging from freshwater to mesohaline (Fig. 1).

The Trout River is a mesohaline tributary of the lower St. Johns River, a 160-km estuarine system that drains ca. 6000 km^2 of urban/industrial, agricultural, and forested lands of northeastern Florida prior to emptying into the Atlantic Ocean (Fig. 2). This entire system has undergone extensive eutrophication and water-quality degradation, largely due to point and non-point source nutrient/toxic chemical loading. Alterations in water-column salinity through tidal exchange and variable freshwater inflows also dramatically influence the chemical/biological nature of the system (Pigg et al., 2004). From May 2001 through May 2003, select meteorological and in situ

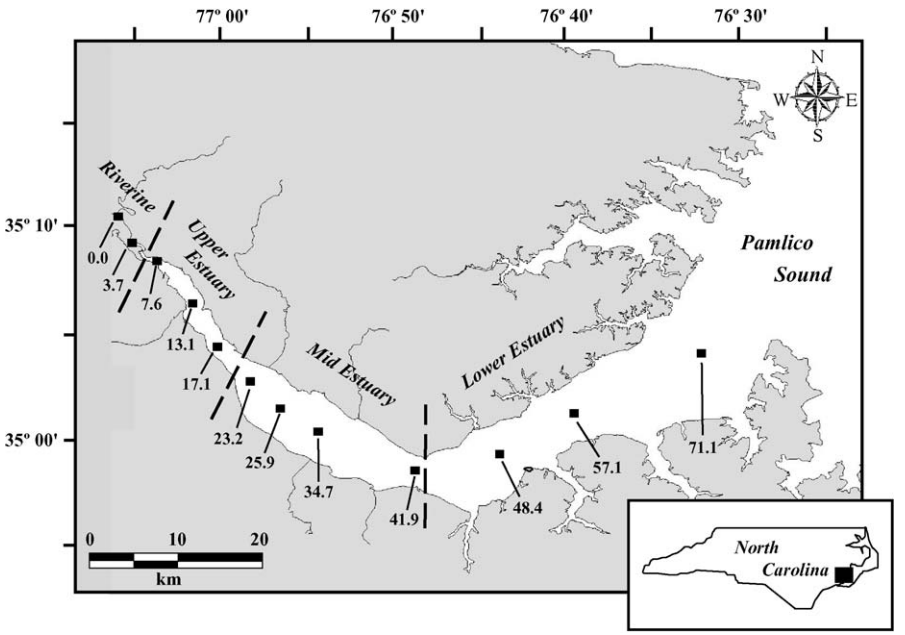

Fig. 1. Location of sampling stations (filled squares) along the length of the Neuse River estuary. Dashed lines differentiate the oligohaline, riverine and upper reach and the mesohaline, mid- and lower-reaches of the estuary (see Section 3). Values for each station represent distance 'downstream' from most 'upstream' site, 0 km. Inset figure places study area relative to North Carolina (USA).

Table 1
Methodology or instrument for acquiring physical, chemical, and biological data via invasive and autonomous sampling in the Neuse and Trout Rivers, respectively

Variable (abbreviation, units)	Neuse River	Trout River
Dissolved oxygen (DO, mg L^{-1})	Hydrolab	YSI 6600
Dissolved oxygen saturation (DO%, %)	Hydrolab	YSI 6600
Conductivity (Cond, mS cm^{-1})	Hydrolab	YSI 6600
Salinity (PSU, practical salinity units)	Hydrolab	YSI 6600
Temperature (Temp, °C)	Hydrolab	YSI 6600, HMP45C
pH (pH, [H^{-1}])	Hydrolab	YSI 6600
Turbidity (Turb, NTU)	–	YSI 6600
Chlorophyll *a* (Chl *a*, μg L^{-1})	Pinckney et al. (1996)	–
Fluorescence (Fluor, as Chl *a* μg L^{-1})	–	WETstar Fluorometer
Wind speed (Wnd Spd, kn h^{-1})	–	MET one windset
Wind direction (Wnd Dir, °)	–	MET one windset
Barometric pressure (BP, mmHg)	–	CS barometric pressure sensor
Precipitation (Precip, mm)	–	RM Young rain gauge
Photosynthetic active radiation (PAR, μmol m^{-2} s^{-1})	Li-COR 4b spherical sensor	Licor quantum sensor
Light attenuation (K_d, m^{-1})	Wetzel (2001)	Wetzel (2001)
Water velocity (Vel, cm s^{-1})	–	Sontek Argonaut-SL
Water depth (Depth, m)	–	Sontek Argonaut
Nitrate + nitrite (NO_x, μg L^{-1})	Jones (1984)	Strickland and Parsons (1972)
Ammonia (NH_4, μg L^{-1})	Solorzano (1969)	Strickland and Parsons (1972)
Phosphate (PO_4, μg L^{-1})	Strickland and Parsons (1972)	Strickland and Parsons (1972)

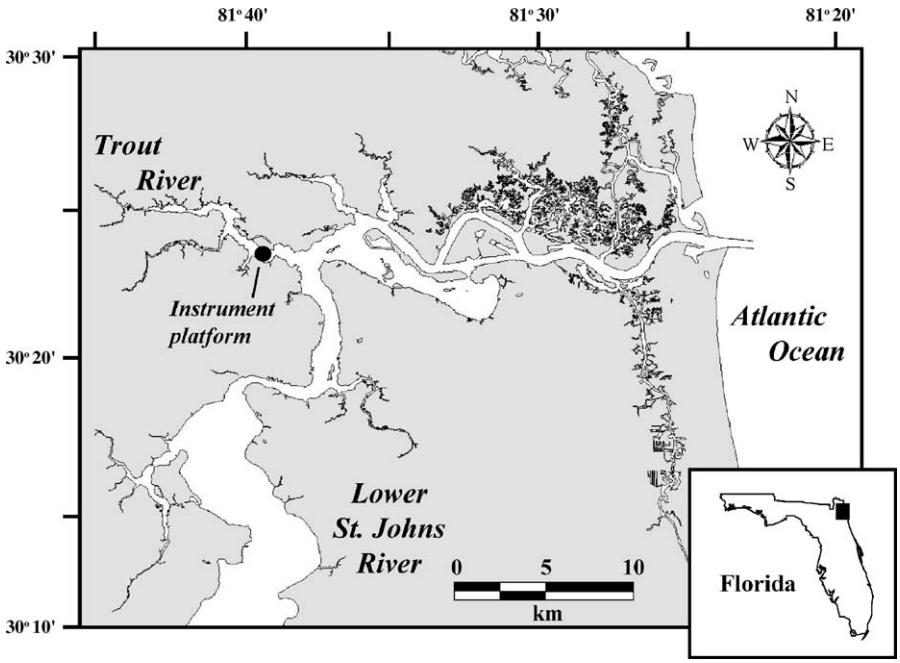

Fig. 2. Location of the autonomous sampling platform (filled circle) within the Trout River, a major tributary of the lower St. Johns River estuarine system. Inset figure places study area relative to northeast Florida (USA).

water-quality data (Table 1) were acquired hourly from mid-channel, sub-surface waters by an autonomous instrument platform (Fig. 2). Large concentrations of dissolved humic materials were present and often dominated water attenuation (see Gallegos, 2002) throughout the lower St. Johns River, including the Trout River. Consequently, accurate Chl *a* concentrations could not always be inferred via in situ fluorescence and fluorescence values were used only to surmise trends of phytoplankton biomass. NO_x, NH_4, and PO_4 concentrations also were determined within sub-surface and bottom waters over 2-week periods during Spring (March/April), Summer (July), and Fall (November/December) from 2000 to 2002 (Table 1). NEP values were calculated from diel, DO concentrations (Caffrey, 2003, 2004). To account, in part, for changing water depth (due to tidal cycles and flow alterations), the mean depth for each sampling interval was used in calculating hourly DO flux. Additionally, the water masses flowing past the sensor were assumed homogeneous (after Caffrey, 2003, 2004). Because NEP values represent the net oxygen flux over a daily interval, diel means of abiotic predictor variables were calculated.

2.2. Statistical analyses

The inherent patterns and trends of abiotic variables and their (interacting) relationships to Chl *a* concentrations and NEP values were determined prior to network development. The associations between/among variables within the Neuse and Trout River were identified using Pearson Product Moment Correlation Coefficients (SYSTAT 10, 2000) to ascertain the adequacy of a variable for model inclusion and/or identify redundant, correlated variables. Data were square root- or logarithmic-transformed (where appropriate) to increase the variance and homogeneity of normalcy. Principal component analysis, utilizing Euclidean distances, characterized sampling sites and dates with respect to physical and chemical variables throughout the Neuse and Trout Rivers (Clarke and Gorley, 2001; Clarke and Warwick, 2001). For annual and seasonal characterization of the Neuse River (based on 30-year temperature regimes; after Litaker, 1986), variable means were calculated as: Winter (December to February), Spring (March to May), Summer (June to September), and Fall (October to November). An analysis of variance (ANOVA; SYSTAT 10, 2000) assessed spatial/temporal differences among mean Chl *a* concentrations.

Because physical/chemical/biological conditions at a 'downstream' estuarine site reflect collective conditions at both that site and 'upstream' sites and algal growth often is temporally 'lagged', exclusion of a temporal and/or spatial sequence within the set of variables may impair the predictive ability of a model (see Duarte, 1990; Pinckney et al., 1997; Maier et al., 1998; Olden, 2000). For that reason, variables were temporally and/or spatially 'lagged' (for the Neuse River, immediately upstream from a site and 2–3 weeks prior to a sampling date and for the Trout River, the previous sampling day), thereby increasing (up to) two-fold the number of potential predictor variables within each data vector. These 'lag' effects were selected to best typify estuarine residence time and/or tidal cycles in the Neuse and Trout Rivers, respectively.

2.2.1. Artificial neural networks

Multi-layer perceptrons using a back-propagation learning algorithm were constructed using NeuroSolutions v4.32 software (NeuroDimension, Inc.; Gainesville, FL, USA) to model Chl *a* concentrations, NEP values, and phototrophic/heterotrophic classifications:

$$\begin{aligned}[\mathrm{Chl}\,a]/\mathrm{NEP} &= f\{W_{P_1,P_3}[f(W_{X_1,P_1}X_1 + W_{X_2,P_1}X_2 + \cdots \\ &\quad + W_{X_i,P_1}X_i + \varepsilon_1)]\} + f\{W_{P_2,P_3}[f(W_{X_1,P_2}X_1 \\ &\quad + W_{X_2,P_2}X_2 + \cdots + W_{X_i,P_2}X_i + \varepsilon_2)]\} \\ &\quad + f\{W_{P_j,P_3}[f(W_{X_1,P_j}X_1 + W_{X_2,P_j}X_2 + \cdots \\ &\quad + W_{X_i,P_j}X_i + \varepsilon_j)]\} \end{aligned} \quad (1)$$

where $X_{1,2,\ldots,i}$ are candidate predictor variables, $P_{1,2,3,\ldots,j}$ are processing elements (PEs), and $W_{X_{1,2,\ldots,i},P_{1,2,\ldots,j}}$ are scalar weights, and $\varepsilon_{1,2\ldots,j}$ is the error (Fig. 3; after Principe et al., 2000).

Briefly, predictor variables were normalized to match the range of the hidden layer's (non-linear) transfer functions (see Goh, 1995; Olden and Jackson, 2002); because hyperbolic tangent functions were used in PEs of the hidden layer for both function

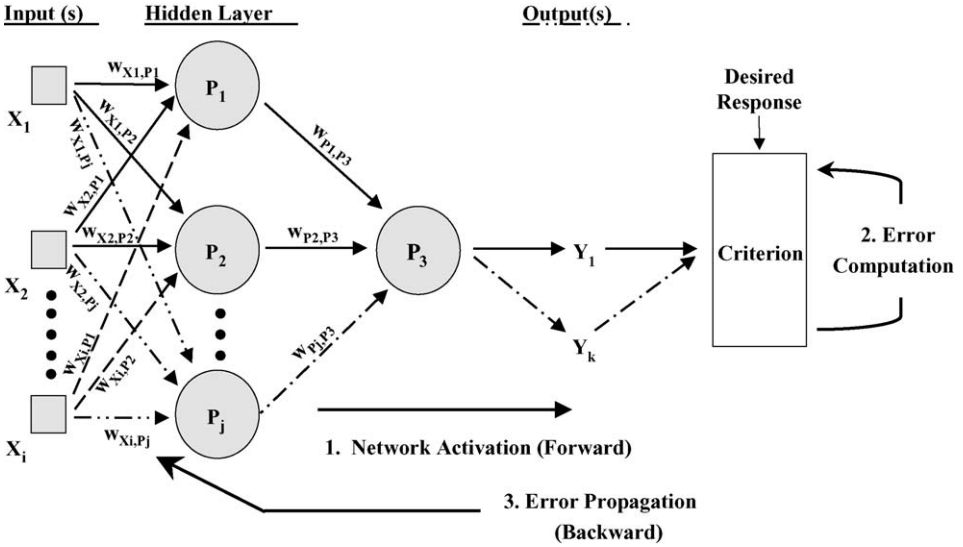

Fig. 3. Schematic of a feed-forward, multi-layer perceptron with back-propagation, depicting the interaction among input variables ($X_{1,\ldots,i}$), processing elements ($P_{1,\ldots,j}$) in the hidden layer(s), synaptic weights ($w_{X_{1,\ldots,i},P_{1,\ldots,j}}$ and $w_{P_{1,\ldots,j},P_3}$), and output parameters ($Y_{1,\ldots,k}$). For each complete presentation of the data set to the network, modeled chlorophyll *a* concentrations or net ecosystem production values were: (1) 'fed forward' for comparison to the desired response, from which (2) mean-square error was computed. The error (3) then was 'back-propagated' and the synaptic weights adjusted (see Section 2).

approximation and classification models, data were scaled from −1 to 1. Normalized values for each data vector (hereafter, an 'exemplar') were multiplied by scalar weights prior to getting summed and processed by multiple transfer functions within the hidden layer. Values generated for hidden-layer PEs then were multiplied by scalar weights prior to getting summed and processed within an output layer to produce a (modeled) output value. Modeled values were scaled to match the chosen transfer function(s) of the output layer; because function approximation and classification models produced output either as a continuous value or a single category, values were scaled to an infinite data range or from 0 to 1, respectively. The modeled value then was 'fed forward' and compared to the desired (measured) response, from which the mean-square error (MSE) was computed. After presentation of all exemplars within a data set (hereafter, an 'epoch'), the error was 'back-propagated' to the network and the weights were incrementally adjusted, through gradient descent with momentum learning, in the direction of the minimum error among PEs (Fig. 3; Principe et al., 2000; Olden, 2000; Lee et al., 2003). In this manner, the weights stabilized over multiple epochs (as error minimized) and modeled values increasingly approximated measured values.

For training, 60% of all exemplars were repeatedly presented to the network (typically 1000–2000 epochs, repeated at least three to five times), with weights adjusted after each epoch to minimize the MSE. To accelerate 'learning' and ensure the greatest probability of network convergence to the global minimum, learning and momentum rates and step-sizes were allowed to vary during iterative training (after Barciela et al., 1999; Principe et al., 2000; Olden, 2000; Olden and Jackson, 2002; Lee et al., 2003). In an attempt to provide an unbiased estimation of a network's predictive success concurrent with training and ensure optimal network design, the MSE also was computed for a 'cross-validation' data sub-set (containing 15% of the exemplars; after Olden, 2000; Olden and Jackson, 2002; Gurbuz et al., 2003). Network training was terminated prior to the designated number of epochs if the MSE within the training or cross-validation data sets fell below 0.01 or began to increase (i.e. an indication that the network began to 'over-train', thereby memorizing the data;

see Karul et al., 2000; Gurbuz et al., 2003). Testing or ''hind-casting'' involved applying a trained network, with frozen weights to a data sub-set (25% of all exemplars) not used in training and cross-validation. Training, cross-validation and testing sub-sets were selected randomly.

Sensitivity analyses and genetic training optimized the choice of input variables and the number/type of input variables, momentum rates/step-sizes, and/or the number of PEs (in the hidden layer), respectively. The resulting models were trained and tested, prior to final selection of the optimal model. Briefly, sensitivity analysis provided an approximate measure of the relative importance among predictor variables by determining the variation of Chl *a* or NEP in response to the variation of individual predictors across a training set. Each predictor variable was varied by a defined number of standard deviations (both +/−) from its mean while all other variables remain fixed (at their respective means; Principe et al., 2000). The Chl *a* concentration or NEP value then was computed (for all deviations) and this process repeated for each variable, after which the most relevant variables (i.e. those creating the greatest variation of the modeled parameter) were identified for network modification.

For genetic training, 50–100 populations of networks were randomly created, trained and evaluated to determine the best fitness (based on the minimum MSE achieved). Attributes of the better-performing networks (i.e. variables, momentum rates, etc.) were combined (using one-point crossover), 'mutated' (using a probability of 0.01), and selected (Roulette-based on Rank) to create a new generation of network populations, which were trained and evaluated. The best attributes again were combined, 'mutated', and selected, creating yet another network generation. In this manner, the better-performing attributes passed along from one 'generation' of networks to the next (typically 100–200 tested), with the optimal-performing network eventually 'evolving' (see Schaffer et al., 1992; Jones, 1993; Montana, 1995).

Network interpretation diagrams (NIDs) illustrated the magnitude/direction of synaptic weights among input-hidden-output layers after training (Aoki and Komatsu, 1999; Chen and Ware, 1999; Özesmi and Özesmi, 1999; Olden, 2000; Olden and Jackson, 2002). Greater weight values indicated more importance in prediction (compared to lesser values) whereas negative values imposed an inhibitory effect (compared to excitatory effect of positive values) on PEs. Because all ANNs used two weight layers (see Fig. 3), positive effects of variables were depicted by positive input-hidden and positive hidden-output weights, or negative input-hidden and negative hidden-output connection weights. Variable interaction was identified by contrasting weights entering the same PE (from Olden, 2000; Olden and Jackson, 2002). The relative share of prediction associated with input variables was determined from final weight values using an algorithm proposed by Garson (1991), and modified by Milne (1995) and Gedeon (1997).

2.3. Regression modeling

To compare results of ANNs with that of linear models, multiple linear regression (MLR) models incorporating identical variables as ANNs and using both non-transformed and transformed values were constructed from training data sets:

$$[\text{Chl}\,a]/\text{NEP} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_i X_i + \varepsilon \quad (2)$$

where $X_{1,\ldots,i}$ are the predictor variables, $\beta_{0,\ldots,i}$ are regression parameters (intercept/slopes of the regression line), and ε is the error (SYSTAT 10, 2000). Regression equations were applied to test data sets, with the (squared) error of data vectors for MLRs and ANNs calculated. A paired *t*-test (SYSTAT 10, 2000) determined whether the mean of the differences for paired MSEs between ANNs and comparable MLRs differed from zero.

3. Results

3.1. Patterns/trends of abiotic/biotic variables

3.1.1. Neuse river estuary

Spatial and temporal groupings of sampling sites, based on seasonal means of physical/chemical variables, were evident. The initial two principal components of the PCA included descriptors indicative of hydrological and meteorological forcing and together explained ca. 81% of the total variability;

PSU, K_d, and NO_x and NH_4 concentrations and Temp and DO concentration explained ca. 52 and 28% of the variability within the first and second PC, respectively. From this ordination, a gradient from the oligohaline riverine- and upper-reaches to the mesohaline mid- and lower-reaches was evident (Fig. 4A). Seasonal groupings of sampling sites also were apparent (Fig. 4B); summer and winter sampling sites were distinct from one another whereas spring and fall sites were relatively similar. PSU and Chl *a* concentrations varied seasonally and along the oligo-/mesohaline gradient, with a pronounced temporal/spatial interaction ($p \leq 0.001$, $n = 2022$). PSU maximized within the lower estuary, with the greatest values occurring during summer (Fig. 4C). Chl *a* concentrations increased along the oligo- to mesohaline gradient, with the least and greatest concentrations occurring within the riverine- and mid-estuary reaches and during summer and winter, respectively (Fig. 4D).

3.1.2. Trout River

Temporal variability among physical/chemical variables occurred, reflecting both estuarine flow and temperature regimes of northeast Florida (Fig. 5A). The initial two PCs of the PCA included descriptors indicative of hydrological and meteorological forcing and together explained ca. 64% of the total variability; PSU and NO_x and PO_4 concentrations, and temperature and NH_4 concentrations explained ca. 37 and 27% of the variability within the first and second PC, respectively. From this ordination, a seasonal continuum was apparent. NEP values from May 2001 to May 2003 indicated that production and respiration processes were variable within the Trout River, resulting in highly dynamic trophic conditions (Fig. 5B). From this, the Trout River appeared phototrophic ca. 66, 46, and 31% of the days during which diel oxygen flux were assessed during 2001 (102 days), 2002 (296 days), and 2003

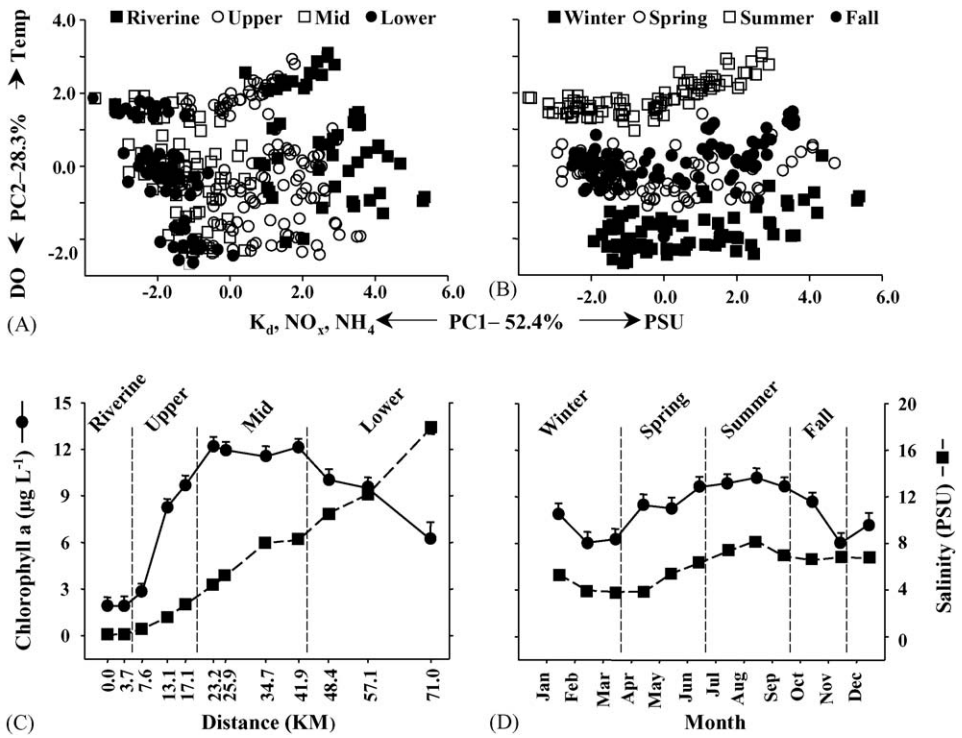

Fig. 4. Two-dimensional principal components (PCs) ordination of Neuse River sampling stations based on physical and chemical parameters. Stations denoted as a function of (A) estuarine reach and (B) season (refer to Fig. 1). Percentages along axes represent percent of total variability explained by the corresponding PC (see Section 3). Chlorophyll *a* concentrations as a function of (C) distance 'downstream' from most 'upstream' site (refer to Fig. 1) and (D) sampling month. Data are means ± standard error ($n = 52$–220).

(141 days), respectively. NEP was highly segmented over short intervals (typically 3–8 days, Fig. 5C).

3.2. Development and validation of ANNs

Numerous multi-layer perceptrons were developed to model Chl *a* concentrations, NEP values, and phototrophic/heterotrophic classifications. Preliminary experimentation with networks possessing PCA and radial-basis function architecture (see Principe et al., 2000) and incorporating varied numbers of PEs and hidden layers, did not improve on results.

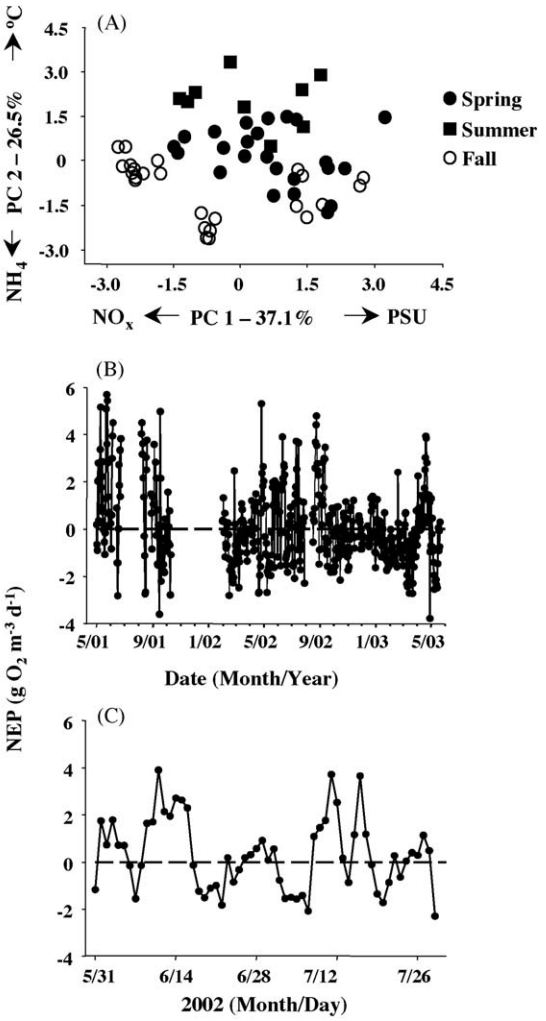

Fig. 5. (A) Two-dimensional principal components (PCs) ordination of Trout River sampling dates based on physical and chemical parameters. Dates denoted as a function of season. Percentages along axes represent percent of total variability explained by the corresponding PC (see Section 3). Values of net ecosystem production (NEP), derived from diel dissolved oxygen concentration, from (B) May 2001 to May 2003 and (C) 31 May to 29 July 2002. Dashed lines within each panel indicate the positive (phototrophic)–negative (heterotrophic) threshold.

3.2.1. Neuse River

An ANN, utilizing the 15 best candidate physical/chemical variables comprised of one-hidden layer with seven PEs, was trained and cross-validated on data sub-sets, prior to being applied to test data. Values of MSE for both the training and cross-validation data sub-sets approached zero (Fig. 6A), indicating that the network had succeeded in 'training' the model. Upon applying the network to a testing data sub-set, modeled concentrations mirrored the general trend in Chl *a* dynamics ($r = 0.64$, $p \leq 0.0001$). However, modeled data dramatically underestimated measured data, particularly at Chl *a* concentrations greater than ca. 20 $\mu g\ L^{-1}$ (Fig. 6B). A sensitivity analysis (Fig. 6C) denoted Temp, PSU, K_d, NO_x, NH_4^+, PO_4^{3-}, and DO to be the most important variables for predicting Chl *a*. An ANN, utilizing these variables and comprised of one-hidden layer with 14 PEs, was successfully trained and cross-validated, with only a slight improvement in modeling Chl *a* concentrations ($r = 0.66$, $p \leq 0.0001$; Fig. 6D). A genetically trained ANN, utilizing the variables, Temp, PSU, K_d, NO_x, L-NO_x, NH_4, L-NH_4, PO_4, DO, and L-Chl *a*, and comprised of one-hidden layer with 11 PEs, did not improve modeling of NEP ($r = 0.67$, $p \leq 0.0001$; data not shown).

3.2.2. Trout River

An ANN, utilizing the best 21 candidate meteorological/physical/chemical variables and comprised of one-hidden layer with four PEs, was trained and cross-validated on data sub-sets, prior to being applied to test data. The network provided a good estimate of modeled NEP values from calculated values ($r = 0.79$, $p \leq 0.0001$; data not shown). A sensitivity analysis (Fig. 7A) denoted the variables, PAR, Depth, Fluor, DO%, Precip, L-PAR, L-Depth, L-Temp, L-Fluor, L-DO%, and L-NEP, to be most important for predicting NEP. An ANN, comprised of one-hidden layer with six PEs and utilizing these 11 variables, was successfully trained and cross-validated, with only a slight improvement in modeling NEP values ($r = 0.82$,

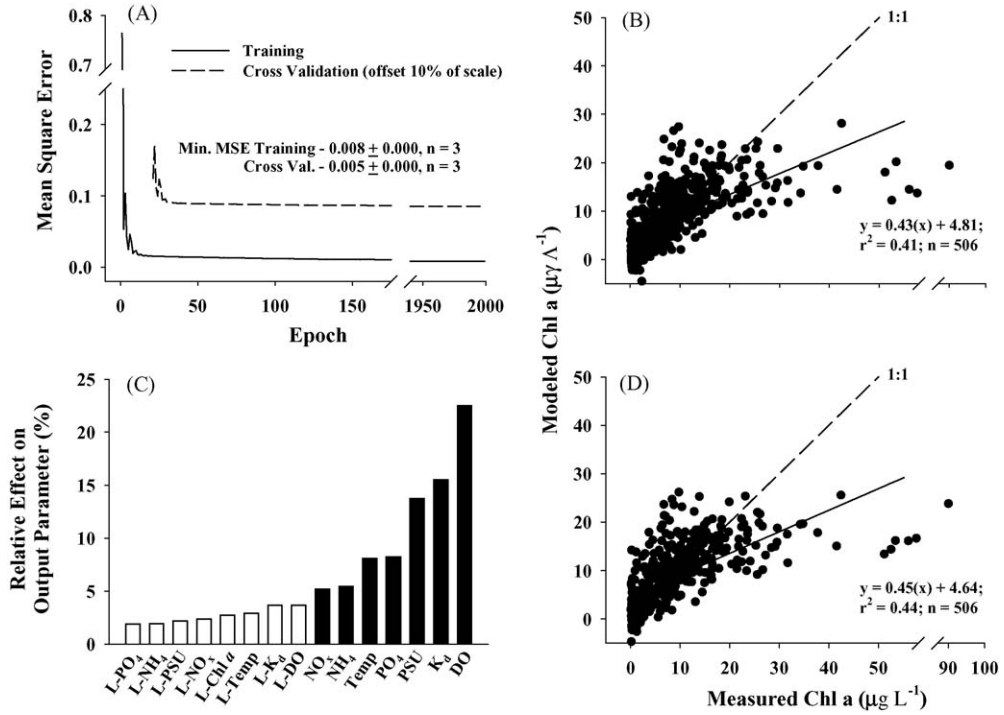

Fig. 6. (A) Mean-square error associated with training and cross-validation data sub-sets during training of an artificial neural network (ANN) incorporating all candidate variables for the Neuse River (see Section 3). Data are means, $n = 3$. (B) Modeled chlorophyll (Chl) *a* concentrations from an ANN incorporating all candidate variables as a function of measured concentrations. The dashed line represents a 1:1 relationship. The solid line and corresponding statistical information represent the 'best' fit relationship, as derived from multiple linear regression. (C) Results of a sensitivity analysis performed on the training data sub-set for the ANN incorporating all candidate variables. Refer to Table 1 for variable abbreviations. 'L' indicates a 'lagged' variable (see Section 2). Filled bars indicate variables selected for subsequent modeling. (D) Modeled Chl *a* concentrations from an ANN using variables selected by a sensitivity analysis as a function of measured concentrations (refer to Figs. 8 and 10A). Lines and statistical information as in (B).

$p \leq 0.0001$; Fig. 7B). A genetically trained ANN, identifying nine predictor variables (°C, PSU, K_d, NO_x, L-NO_x, NH_4^+, L-NH_4, PO_4^{3-}, DO, L-Fluor) and comprised of one hidden layer with seven PEs, did not improve modeling of NEP values ($r = 0.78$, $p \leq 0.0001$; Fig. 7C). In all networks, modeled NEP values slightly overestimated and underestimated measured values at the least and greatest NEP values.

ANNs predicted instances of phototrophy or heterotrophy fairly well, with classification errors of ca. 22% and Cohen's Kappa values of ca. 0.55 ± 0.01 S.E. [*Note*: the kappa statistic denoted the degree of concurrence between calculated-modeled sorting of trophic classifications; values greater than 0.75 indicated strong agreement whereas values between 0.40 and 0.79 indicate fair to good agreement; SYSTAT 10, 2000]. An ANN, utilizing 10 input variables (Depth, L-Depth, Fluor, L-Fluor, DO, L-DO, L-Precip, L-PAR, L-Temp, and L-Turb) and 4 PEs developed via sensitivity analysis, predicted the proper classification for ca. 69 and 85% of phototrophic and heterotrophic instances, respectively. A genetically trained ANN, utilizing 13 inputs (Wnd Dir, L-Wnd Dir, PAR, Depth, PSU, Turb, Fluor, DO%, L-DO%, L-Wnd Spd, L-Depth, L-Temp, and L-Heterotrophic class) and 2 PEs, produced similar results (Table 2).

3.3. NIDs and Garson's algorithms

Based on final weight values, NIDs of the best predictive networks for the Neuse and Trout Rivers (Figs. 8 and 9, respectively) indicated extremely complex interactions among abiotic and biotic

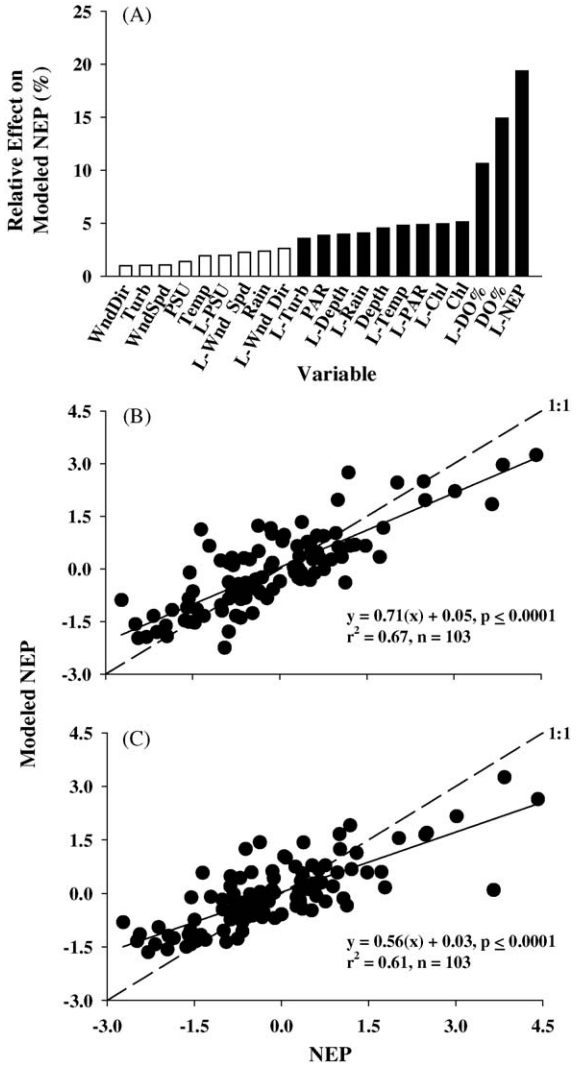

Fig. 7. (A) Results of a sensitivity analysis performed on the training data sub-set from the artificial neural network (ANN) incorporating all candidate variables for the Trout River. Refer to Table 1 for variable abbreviations. 'L' indicates a 'lagged' variable (see Section 2). Filled bars indicate variables selected for subsequent modeling. (B) Modeled net ecosystem production (NEP) values from an ANN using variables selected by a sensitivity analysis as a function of actual values (refer to Figs. 9 and 11A). (C) Modeled NEP values (B) from an ANN derived using genetic training as a function of measured values (refer to Fig. 11B). Dashed lines represent a 1:1 relationship. Solid lines and corresponding statistical information represent the 'best' fit relationship, as derived from multiple linear regression.

variables. No variable in either network had a consistent magnitude or direction (positive or negative) of impact among PEs within the input-hidden layers. Rather, the apparent strong positive influences of PSU and DO and PAR, L-Temp, L-DO and L-NEP on select PEs for predicting Chl *a* and NEP, respectively, were 'counter-balanced' by equal or lesser negative influences of the same variables among alternative PEs.

Based on absolute network weights of derived from sensitivity analysis and genetic training for the Neuse River, Garson's algorithm denoted PSU, DO, and nutrient concentrations to have the greatest relative impact on prediction of Chl *a* concentrations (Fig. 10A and B). For the ANNs derived from sensitivity analysis and genetic training for the Trout River, Garson's algorithm denoted both contemporary and 'lagged' variables (Fluor, DO%, Turb, L-Turb, and L-NEP) to have the greatest impact on prediction of NEP values (Fig. 11A and B). No particular one (or few) variable(s) proved to greatly influence Chl *a* or NEP to the exclusion of other variables; rather multiple variables in both the Neuse and Trout Rivers had ca. similar relative effects.

3.4. Comparison of ANNs and MLR

The MSE for all trained ANNs was equal to or less than that for comparable MLR models, indicating that networks performed as well as, or outperformed all linear models (Table 3). MLR models using transformed variables did not always outperform MLR models using non-transformed variables.

4. Discussion

An intuitive premise adopted in these modeling efforts was that the interplay of physical/chemical variables created environmental gradients responsible for the transitory and spatially explicit patterns of phytoplankton abundance and system-level production (e.g. Dustan and Pinckney, 1989; Pinckney and Dustan, 1990; Cloern, 1991; Klarer and Millie, 1994; McKee et al., 2002; Millie et al., 2003, 2004; Pigg et al., 2004). Spatial and temporal trends of sampling sites and/or dates were evident, with differences attributable to suites of variables arising from fresh-

Table 2
Confusion matrix, delineating the number of modeled phototrophic and heterotrophic classifications from ANNs derived using sensitivity analysis and genetic training with the calculated classifications for the Trout River (refer to Fig. 4A and B)

ANN architecture	Calculated classification	Modeled phototrophic	Modeled heterotrophic
Ten variables, four processing elements derived via sensitivity analysis	Phototrophic	34 (69.4)	15
	Heterotrophic	8	45 (85.2)
Thirteen variables, two processing elements derived via genetic training	Phototrophic	40 (81.6)	9
	Heterotrophic	15	39 (72.2)

Numbers in parentheses signify the percentage of correct classifications. See Section 3 for listing of predictor variables and associated Kappa statistics.

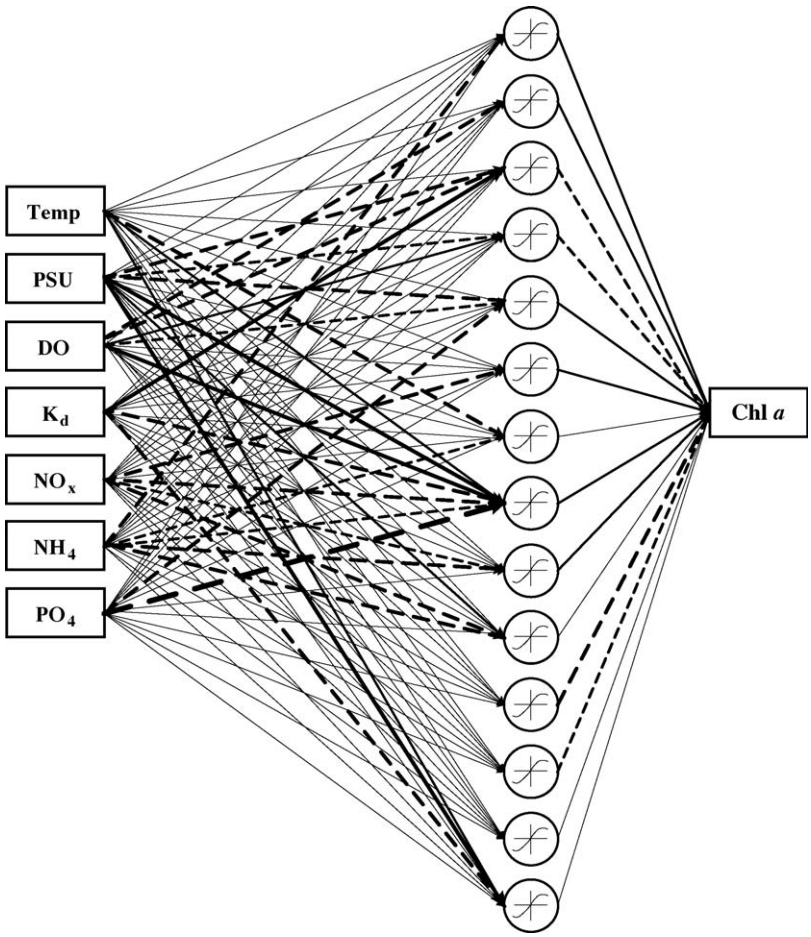

Fig. 8. An artificial neural network interpretation diagram (7 input variables, 14 processing elements) for hindcasting chlorophyll (Chl) *a* concentrations in the Neuse River. Dashed and solid lines depict negative (inhibitory) or positive (excitatory) effects, respectively, upon modeled Chl *a* concentrations by synaptic weights among input-hidden-output layers. Line thickness portrays the relative magnitude of the weight. Refer to Table 1 for variable abbreviations. 'L' indicates a 'lagged' variable (see Section 2).

water inflows, saltwater influx/tidal influences, and/or meteorological conditions. Chl *a* mirrored system-level salinity regimes in the Neuse River; mean concentrations were greatest in the mid- and lower estuary where lesser turbidity, current velocities, and nutrient loads, and greater salinity and residence times (compared to the upper reaches) exist. Mean concentrations were least during cool, winter months and greatest during warm, summer months. NEP values in the Trout River were extremely dynamic with continuous phototrophic or heterotrophic conditions often occurring over short-time intervals (days). This alternate 'source and sink' for carbon is typical for coastal estuaries (see Smith and Hollibaugh, 1993; Caffrey, 2003, 2004) and most likely resulted from varying tidal cycles coupled with episodic occurrences of hydrologic discharge, water-column salinity stratification, phytoplankton bloom events, and meteorological fronts.

ANNs developed for the Trout River reliably modeled NEP values and trophic classifications. ANNs developed for the Neuse River performed adequately at Chl *a* concentrations less than 20 $\mu g\ L^{-1}$. Chl *a* concentrations greater than 20 $\mu g\ L^{-1}$ were dramatically underestimated by all networks, likely due to one or more reasons. Foremost, the periodicity of data acquisition may not have been sufficient to generalize system-level variability. Phytoplankton growth in coastal waters is governed by interacting processes (e.g. nutrient uptake, temperature dependence, light availability, grazing impacts; e.g. Duarte, 1990) driven by multiple, system-level factors (i.e., meteorological conditions, hydrologic regime, intermittent water-column gradients, benthic resuspension; e.g. Paerl,

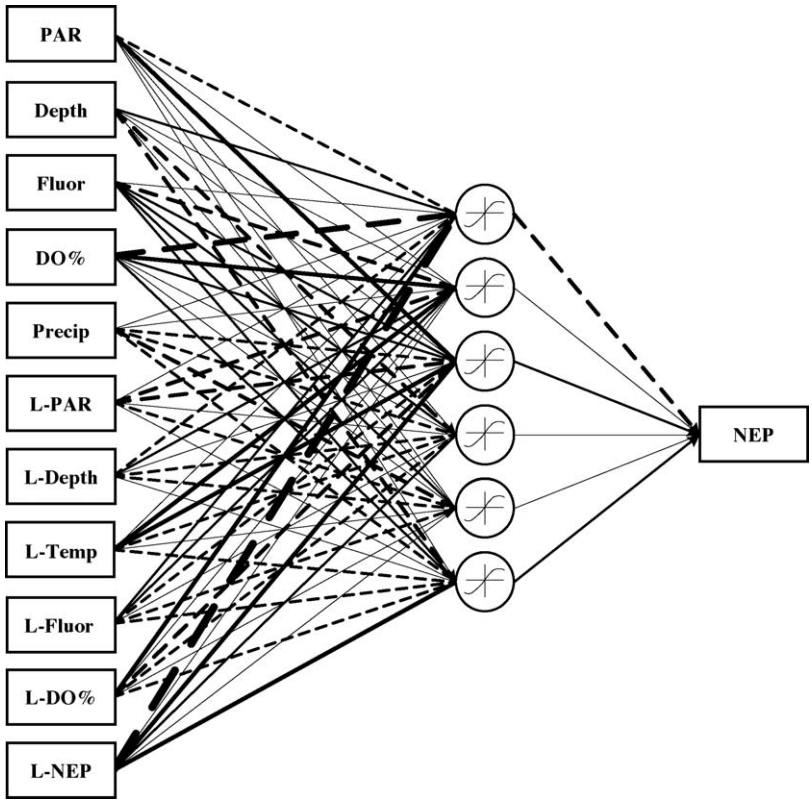

Fig. 9. An artificial neural network interpretation diagram (11 input variables, 6 processing elements) for hindcasting net ecosystem production (NEP) values in the Trout River. Dashed and solid lines illustrate negative (inhibitory) or positive (excitatory) effects, respectively, upon modeled NEP values by synaptic weights among input-hidden-output layers. Line thickness portrays the relative magnitude of the weight. Refer to Table 1 for variable abbreviations. 'L' indicates a 'lagged' variable (see Section 2).

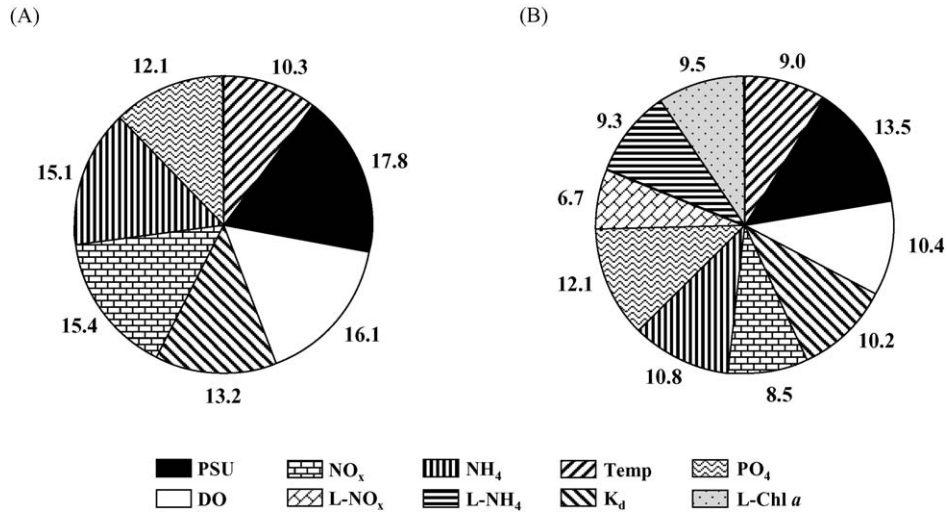

Fig. 10. The relative share of prediction associated with abiotic/biotic variables in hindcasting chlorophyll *a* concentrations in the Neuse River with an artificial neural network incorporating variables derived using (A) sensitivity analysis and (B) genetic training. Refer to Table 1 for variable abbreviations. 'L' indicates a 'lagged' variable (see Section 2).

1988; Paerl et al., 1998, 1999; Millie et al., 2003, 2004), often interacting on short (day to week) time scales. Consequently, the sampling interval in the Neuse River (generally 2–3 weeks) likely did not provide appropriate resolution for the ANNs to capture the inherent variability and magnitude of abiotic variables and/or Chl *a* dynamics throughout the system. Lee et al. (2003) noted that a minimum

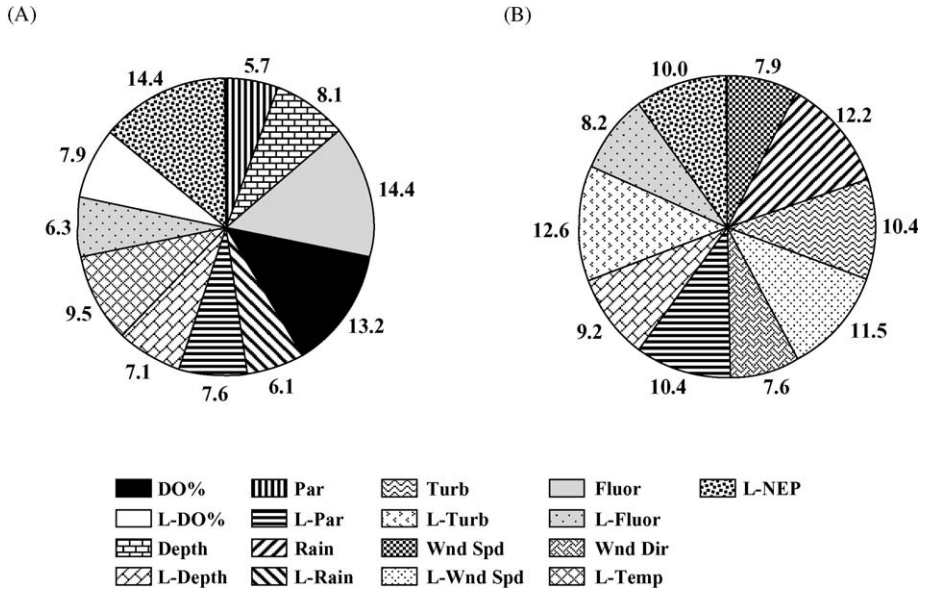

Fig. 11. The relative share of prediction associated with abiotic/biotic variables in hindcasting net ecosystem production values in the Trout River with an artificial neural network incorporating variables derived using (A) sensitivity analysis and (B) genetic training. Refer to Table 1 for variable abbreviations. 'L' indicates a 'lagged' variable (see Section 2).

Table 3
Mean square error derived from distinct model formulations, utilizing non-transformed and transformed multiple linear regression (MLR and TR-MLR, respectively) and artificial neural network (ANN) architectures, for hindcasting chlorophyll *a* (Neuse River) and net ecosystem production (Trout River)

Model formulation	Model architecture	Chlorophyll *a*	Net ecosystem production
All candidate variables	MLR	59.35 ± 14.80 (=0.066)	0.72 ± 0.12 (=0.558)
	TR-MLR	69.46 ± 15.43 (≤0.001)	0.69 ± 0.11 (=0.692)
	ANN	53.16 ± 12.14	0.66 ± 0.10
Inputs derived via sensitivity analysis	MLR	59.82 ± 14.14 (=0.014)	0.62 ± 0.11 (=0.642)
	TR-MLR	71.37 ± 16.42 (≤0.001)	0.63 ± 0.11 (=0.446)
	ANN	50.38 ± 11.14	0.59 ± 0.09
Inputs derived via genetic algorithm	MLR	58.26 ± 13.99 (=0.074)	0.68 ± 0.11 (=0.769)
	TR-MLR	68.48 ± 15.40 (=0.004)	0.69 ± 0.11 (=0.801)
	ANN	49.74 ± 10.82	0.71 ± 0.14

Refer to Section 3 for input variables of distinct model formulations. Data are means ± standard error ($n = 104$). The significance of the difference between paired data vectors for regressions and network architectures is in parentheses (see Section 2).

sampling interval of 1 week was necessary to produce a data set from which an ANN could accurately reproduce phytoplankton dynamics.

Moreover, the designated 'lag' effects may not have accurately portrayed the spatial and temporal separation of 'cause and effect' in phytoplankton growth and accumulation throughout the Neuse River, thereby decreasing the predictive capacity of the models. 'Lag' effects (for the Neuse River, immediately upstream from a site and 2–3 weeks prior to a sampling date and for the Trout River, the previous sampling day) were selected to best typify estuarine residence time and/or account for the impacts of variable flows and daily tidal cycles. For networks predicting Chl *a* concentrations in the Neuse River, sensitivity analysis and genetic training selected contemporary variables and 'lagged' and contemporary variables, respectively. Nevertheless, selected variables were indicative of riverine flow (e.g. PSU, nutrient concentrations, K_d, etc.) and/or proxy measurements for biomass (DO), indicating the importance of hydrological forcing in regulating phytoplankton accumulation throughout the estuary (see Rudek et al., 1991; Mallin et al., 1993; Pinckney et al., 1997, 1999). Interestingly, variables known to affect phytoplankton production (PAR, Temp, Turb), and/or act as proxy measurements for biomass (Fluor) comprised the predictor variables for NEP in the Trout River. Approximately one-half of the variables selected to predict NEP in the Trout River were 'lagged' variables, thereby implying a non-linear, autoregressive process for system trophic state (after Lee et al., 2003). For ANNs modeling the onset and magnitude of cyanobacterial blooms within the River Murray (South Australia), Maier et al. (1998) concluded that 'lagged' water-quality variables were vital, and often superior predictors to contemporary variables. Increasing the spatial and temporal 'lag effects' (to include variables farther upstream and/or over a greater temporal period) in models for both estuaries neither improved network prediction nor altered model interpretation.

Better performance of a network using the Neuse River data set may not have been realistic. Phytoplankton blooms throughout the Neuse River typically were infrequent and spatially distinct, often restricted to the mid and lower reaches during low flow periods when nutrient accumulation with a stagnated water-column fuels algal growth and proliferation (Mallin et al., 1993; Mallin, 1994; Pinckney et al., 1997, 1999). Consequently, the data sub-set with which the ANNs were trained contained few Chl *a* concentrations greater than 25 $\mu g\,L^{-1}$. To investigate the (overall) variation in the data, a large network (20 PEs) was trained on the entire data set for a time to sufficiently allow data memorization (5000 epochs). Although this network was not appropriate for prediction, it did present a means to assess how well the 'best possible' ANN for a particular data set performed across selected data sub-sets. Greater correspondence between measured and modeled concentrations occurred when this network was

applied to data sub-sets for the riverine and upper reaches ($r = 0.64$, $n = 64$, $p \leq 0.0001$ and $r = 0.71$, $n = 605$, $p \leq 0.0001$, respectively) than for the mid and lower reaches ($r = 0.59$, $n = 743$, $p \leq 0.0001$ and $r = 0.55$, $n = 311$, $p \leq 0.0001$, respectively; data not shown). Clearly, considerable variation in Chl *a* existed (more so in the lower estuary where bloom conditions were most prevalent) that could not be correlated to the selected abiotic variables. Given the data set, there appeared to be a predictive 'limit' for modeling Chl *a* with a network.

Although there is uncertainty as to whether an untested variation in network parameters (i.e. number/value of hidden nodes, PEs, learning rates, etc.) might not have provided better prediction (Garson, 1991), abiotic variables characterized differences among sampling sites and dates and served as predictors for the trends/patterns of both Chl *a* and NEP Only physical/chemical variables routinely collected in or analyzed for by invasive or autonomous-based sampling programs in the Neuse and Trout Rivers, respectively, were chosen for potential inclusion in the networks. Obviously, one (or several) variable(s) that would have dramatically improved network prediction might have been excluded. Intuitively, inclusion of variables that can act as proxy measurements for phytoplankton abundance would increase the predictive capability for Chl *a*. For example, particulate organic carbon concentrations and/or carbon:nitrogen ratios were available within the Neuse River data set; however, these variables were directly correlated with phytoplankton biomass and excluded from initial network development (see Section 2). When these variables were introduced as predictors in a network, the correspondence ($r = 0.82$, $p \leq 0.0001$) between modeled and measured Chl *a* concentrations, along with the predictive capacity of the network ($r^2 = 0.68$, $p \leq 0.001$) was dramatically greater than previous ANNs. However, the absence of protocols for routine (and inexpensive) acquisition of carbon-based variables prohibits their use as potential predictor variables (either in addition to or to the exclusion of other more readily obtainable abiotic variables) in estuarine modeling efforts.

ANNs proved an attractive substitute (from MLR) for modeling the large, complex data sets of the Neuse and Trout Rivers. Regression analysis often is used for predicting algal biomass based on its direct correspondence with a single (or multiple) abiotic/biotic variable(s) (e.g. Sarnelle, 1992; Brown et al., 2000; Bachmann et al., 1996, 2001). Networks generally outperformed MLR models (based on minimization of MSE), denoting the apparent non-linear and/or stochastic interactions among abiotic/biotic variables throughout both estuaries. In theory, an ANN encompasses the MLR model; as such, networks should perform as well, or better, than linear models for such complex systems due to their inherent flexibility in dealing with the non-linear influences of multiple variables (Scardi and Harding, 1999; Gonzalez, 2000). This greater performance is not surprising in a strict statistical sense; ecological data often are not normally distributed and even if data are transformed, MLR is not particularly successful in modeling such data (Maier et al., 1998).

A requirement for linear regression is an a priori knowledge of appropriate predictors for model inclusion or exclusion. In the absence of knowledge concerning the underlying relationships among estuarine abiotic and biotic variables, step-wise MLR may result in models composed of variables having no theoretical relationship to Chl *a* and NEP. However, we should not focus on whether one model can outperform another model, but rather, address the relative performance of the models in the absence of known relationships (Smith and Mason, 1997). Herein lies the dilemma of modeling meta-stable dynamics in self-organizing systems (such as dynamic coastal waters) that are created and stabilized through internal interactions among scales (Perry, 1995); if predictable, linear relationships between physical/chemical variables and phytoplankton assemblages existed, modeling and forecasting estuarine biomass and production would be straightforward (see Cloern, 2001).

5. Concluding remarks

The means to integrate accurate prediction with interpretable system-level information is a basic tenet of ecological forecasting. Clearly, accurate characterization of environmental variables used to capture the timing and magnitude of biotic indicators is required. Although it often is difficult to predict 'indicator outcomes' within complex aquatic systems exhibiting

vertical/horizontal variances across spatial/temporal scales, statistical-based models can reveal consistent 'large-number' relationships (e.g. the association between phosphorus and phytoplankton; Harris, 1994). Moreover, estuarine systems "... are sufficiently close to the 'edge of chaos' that repeatable system-level properties ... do frequently emerge ...", and as a consequence, "... a predictable system of self-organizing components may therefore arise ..." (Harris, 1996).

Abiotic variables served as (interacting) 'small number' predictors for the dynamic and often, stochastic trends of 'large number' indicators (Chl *a* concentrations, NEP values, and trophic classifications) within the Neuse and Trout Rivers. Few variables proved to greatly impact Chl *a* and NEP; rather, suites of interacting physical/chemical variables indicative of meteorological and hydrological forcing and/or proxy measurements of phytoplankton accumulation were selected as the best predictor variables. The 'knowledge' of an ANN is contained in the synaptic weights and interpreting (trained) network 'equations' is both intrinsically difficult and non-sensical (refer to Eq. (1) and Figs. 8 and 9); as such, quantitative information concerning the absolute impact of and/or relationships among predictor variables could not be easily obtained. Although sensitivity analysis and Garson's algorithm deconvolved the relative impact of variables in predicting Chl *a* and NEP, the positive/negative direction of the variables was not taken into account within calculations of these analyses. Consequently, interpretation was limited to the overall magnitude of the (relative) effect of predictor variables (see Olden and Jackson, 2002).

Effective forecasting of the (chronic and episodic) impacts of stressors and/or disturbances throughout estuarine waters necessitates robust, data-assimilative modeling approaches. Vast amounts of spatial and temporal-intensive data are being collected within coastal monitoring programs (often through autonomous sampling and analysis protocols) for the establishment of 'baseline' conditions against which to gauge ecological change (Paerl et al., 2005). Statistical models generally are not reliable outside the range of data used for development (Maier et al., 1998; Karul et al., 2000). Although network development for both systems encompassed multi-year data sets, greater data variability than that observed here is common for estuaries located in the southeastern USA (e.g. during times of episodic, wide-range disturbance, Paerl et al., 2001; Tester et al., 2003). Site-specific ANNs incorporating such high-resolution and temporally variable data would have greater forecasting capability than regional and/or universal models encompassing systems exhibiting lesser variability and/or less-diverse sampling/analytical protocols. Clearly, site-specific networks (such as that presented here) need to be continually 'updated' for realistic characterization of biotic indicators along ecologically relevant scales and pertinent to requirements of synoptic water-resource management (after Millie et al., 1995).

ANNs are considered by many scientists to be 'black boxes' (see Garson, 1991; Smith and Mason, 1997; Olden and Jackson, 2002), and except for their predictive capability, might appear to be of limited value for ecological applications and problem solving. The crucial 'next step' for routine utilization of ANNs in modeling relevant biotic indicators and/or as a means to discern estuarine function is the utilization of statistical approaches or 'rule extraction' algorithms (e.g. 'if–then' input intervals, 'validity interval analysis', 'data mining', etc.; Quinlan, 1986; Andrews et al., 1995; Thrun, 1995; Craven and Shavlik, 1996) to allow for comprehensible network interpretation. For example, Olden and Jackson (2002) utilized a randomization approach to statistically assess the importance of synaptic weights and the contribution of environmental variables to an ANN modeling fish species richness within lakes. Soyupak and Chen (2004) utilized a fuzzy-logic model (based on heuristic knowledge, rather than the input–output relations of an ANN) to approximate the functional non-linear relationships between Chl *a* concentrations and water-quality variables within a reservoir. Such diverse approaches can improve predictive capability while simultaneously identifying non-additive structure in the data and simplifying the holistic interpretation of interacting variables (Millie et al., 1995).

Acknowledgements

This work was funded, in part, by grants from the National Oceanic & Atmospheric Administration—Coastal Monitoring & Assessment Program, the

National Science Foundation, the U.S. Department of Agriculture—NRI Program, the U.S. Environmental Protection Agency—STAR Program, and the North Carolina Sea Grant Program. We express appreciation to C. Donahue, M. Fitzpatrick, M. Go, K. Howe, B. Peierls, M. Piehler, J. Swistack, C. Talent, R. Weaver, and P. Wyrick assisting in field sampling and analytical/technical support. S. Cook assisted in figure preparation. Reference to proprietary names are necessary to report factually on available data; however, the Florida Institute of Oceanography, Ohio University, University of North Carolina-Chapel Hill, University of South Carolina, the Florida Fish & Wildlife Conservation Commission, and the National Oceanic & Atmospheric Administration neither guarantee nor warrant the standard of a product and imply no approval of a product to the exclusion of others that may be suitable.

References

Andrews, R., Diederich, J., Tickle, A.B., 1995. A survey and critique of techniques for extracting rules from trained artificial neural networks. Knowl.-Based Syst. 8, 373–389.

Aoki, I., Komatsu, T., 1999. Analysis and prediction of the fluctuation of sardine abundance using a neural network. Oceanol. Acta 20, 81–88.

Bachmann, R.W., Hoyer, M.V., Canfield Jr., D.E., 2001. Evaluation of recent limnological changes at Lake Apopka. Hydrobiologia 448, 19–26.

Bachmann, R.W., Jones, B.L., Fox, D.D., Hoyer, M., Bull, L.A., Canfield Jr., D.E., 1996. Relations between trophic state indicators and fish in Florida (USA) lakes. Can. J. Fish. Aquat. Sci. 53, 842–855.

Barciela, R.M., Garcia, E., Fernández, E., 1999. Modelling primary production in a coastal embayment affected by upwelling using ecosystem models and artificial neural networks. Ecol. Model. 120, 199–211.

Barnes, T., Mazzotti, F.J., 2005. Using conceptual models to select ecological indicators for monitoring, restoration, and management of estuarine ecosystems. In: Bortone, S.A. (Ed.), Estuarine Indicators. CRC Press, Boca Raton, FL, pp. 493–501.

Bortone, S.A., 2005. The quest for the ''perfect'' estuarine indicator: an introduction. In: Bortone, S.A. (Ed.), Estuarine Indicators. CRC Press, Boca Raton, FL, pp. 1–3.

Brown, C.D., Hoyer, M.V., Bachmann, R.W., Canfield Jr., D.E., 2000. Nutrient–chlorophyll relationships: an evaluation of empirical nutrient–chlorophyll models using Florida and north-temperate lake data. Can. J. Fish. Aquat. Sci. 57, 1574–1583.

Caffrey, J.M., 2003. Production, respiration, and net ecosystem metabolism in U.S. estuaries. Environ. Monit. Assess. 81, 207–219.

Caffrey, J.M., 2004. Factors controlling net ecosystem metabolism in U.S. estuaries. Estuaries 27, 90–101.

Caffrey, J.M., Cloern, J.E., Grenz, C., 1998. Changes in production and respiration during a spring phytoplankton bloom in San Francisco Bay, California USA: implications for net ecosystem metabolism. Mar. Ecol. Prog. Ser. 172, 1–12.

Chen, D.G., Ware, D.M., 1999. A neural network model for forecasting fish stock recruitment. Can. J. Fish. Aquat. Sci. 56, 2385–2396.

Clarke, K.R., Gorley, R.N., 2001. PRIMER v5: User Manual/Tutorial. Primer-E, Plymouth, UK, 91 pp.

Clarke, K.R., Warwick, R.M., 2001. Change in Marine Communities: An Approach to Statistical Analyses and Interpretation, 2nd ed. Primer-E, Plymouth, UK.

Cloern, J.E., 1991. Tidal stirring and phytoplankton bloom dynamics in an estuary. J. Mar. Res. 49, 203–221.

Cloern, J.E., 2001. Our evolving conceptual model of the coastal eutrophication problem. Mar. Ecol. Prog. Ser. 210, 223–253.

Committee on Environmental and Natural Resources, 1997. Integrating the nation's environmental and research networks and programs: a proposed framework. National Science & Technology Council special Report. Office of Science & Technology Policy, Washington, DC, 103 pp.

Cooper, S.R., McGlothlin, S.K., Madritch, M., Jones, D.L., 2004. Paleoecological evidence of human impacts on the Neuse and Pamlico estuaries of North Carolina, USA. Estuaries 27, 617–633.

Craven, M.W., Shavlik, J.W., 1996. Extracting tree-structured representations of trained networks. In: Touretzky, D., Mozer, M., Hasselmo, M. (Eds.), Advances in Neural Information Processing Systems 8. Massachusetts Institute of Technology Press, Cambridge, MA, pp. 24–30.

Duarte, C.M., 1990. Time lags in algal growth: generality, causes and consequences. J. Plankton Res. 12, 873–883.

Dustan, P., Pinckney, J., 1989. Tidally induced estuarine phytoplankton patchiness. Limnol. Oceanogr. 34, 408–417.

Gallegos, C.L., 2002. An optical water quality model for the lower St Johns River. In: Proceedings of the Lower St. Johns River Research Coordination Conference, St. Johns River Water Management District Palatka, Florida, October, pp. 40–45.

Garson, G.D., 1991. Interpreting neural-network connection weights. Artif. Intell. Expert 6, 47–51.

Gedeon, T.D., 1997. Data mining of inputs: analyzing magnitude and functional measures. Int. J. Neural Syst. 8, 209–218.

Glibert, P., Conley, D., Fisher, T., Harding, L., Malone, T., 1995. Dynamics of the 1990 winter/spring bloom in Chesapeake Bay. Mar. Ecol. Prog. Ser. 122, 27–43.

Goh, A.T.C., 1995. Back-propagation neural networks for modeling complex systems. Artif. Intell. Eng. 9, 143–151.

Gonzalez, S., 2000. Neural Networks for macroeconomic forecasting: a complimentary approach to linear regression models. Canadian Department of Finance Working Paper No. 2000-07, Ottawa, Ont., Canada, 40 pp.

Gurbuz, H., Kirvak, E., Soyupak, S., Yerli, S., 2003. Predicting dominant phytoplankton quantities in a reservoir by using neural networks. Hydrobiologia 504, 133–141.

Harris, G.P., 1994. Pattern, process and prediction in aquatic ecology—a limnological view of some general ecological problems. Freshwater Biol. 15, 261–266.

Harris, G.P., 1996. A reply to Sarnelle (1996) and some further comments on Harris's (1994) opinions. Freshwater Biol. 35, 343–347.

Jones, A.J., 1993. Genetic algorithms and their applications to the design of neural networks. Neural Comput. Appl. 1, 32–45.

Jones, M., 1984. Nitrate reduction by shaking with cadmium: alternative to cadmium columns. Water Res. 18, 943–946.

Jordon, S.J., Smith, L.M., 2005. Indicators of ecosystem integrity for estuaries. In: Bortone, S.A. (Ed.), Estuarine Indicators. CRC Press, Boca Raton, FL, pp. 467–480.

Karul, C., Soyupak, S., Cilesiz, A., Akbay, N., Germen, E., 2000. Case studies on the use of neural networks in eutrophication modeling. Ecol. Model. 134, 145–152.

Klarer, D.M., Millie, D.F., 1994. Regulation of phytoplankton dynamics in a Laurentian Great Lakes estuary. Hydrobiologia 286, 97–108.

Lee, J.H.W., Huang, Y., Dickman, M., Jayawardena, A.W., 2003. Neural network modeling of coastal algal blooms. Ecol. Model. 159, 179–201.

Litaker, R.W., 1986. Dynamics of a well-mixed estuary. Ph.D. Dissertation. Duke University, Durham, North Carolina, 523 pp.

Maier, H.R., Dandy, G.C., Burch, M.D., 1998. Use of artificial neural networks for modeling cyanobacteria *Anabaena* spp. in the River Murray, South Australia. Ecol. Model. 105, 257–272.

Mallin, M., 1994. Phytoplankton ecology of North Carolina estuaries. Estuaries 17, 561–574.

Mallin, M.A., Paerl, H.W., Rudek, J., Bates, P.W., 1993. Regulation of estuarine primary production by watershed rainfall and river flow. Mar. Ecol. Prog. Ser. 93, 199–203.

Marshall III, F.E., 2005. Using statistical models to simulate salinity variability in estuaries. In: Bortone, S.A. (Ed.), Estuarine Indicators. CRC Press, Boca Raton, FL, pp. 33–52.

Mazumder, A., 1994. Patterns of algal biomass in dominant odd- vs. even-link lake ecosystems. Ecology 75, 1141–1149.

McKee, D., Cunningham, A., Jones, K.J., 2002. Optical and hydrographic consequences of freshwater run-off during spring phytoplankton growth in a Scottish fjord. J. Plankton Res. 24, 1163–1171.

Millie, D.F., Paerl, H.W., Hurley, J.P., 1993. Microalgal pigment assessments using high-performance liquid chromatography: a synopsis of organismal and ecological applications. Can. J. Fish. Aquat. Sci. 50, 2513–2527.

Millie, D.F., Carrick, H.J., Doerong, P.H., Steidinger, K.A., 2004. Intra-annual variability of water quality and phytoplankton within the St. Lucie River Estuary (Florida, USA): a quantitative perspective. Est. Coast. Shelf Sci. 61, 137–149.

Millie, D.F., Fahnenstiel, G.L., Lohrenz, S.E., Carrick, H.J., Johengen, T.H., Schofield, O.M.E., 2003. Physical–biological coupling in southern Lake Michigan: influence of episodic resuspension on phytoplankton. Aquat. Ecol. 37, 393–408.

Millie, D.F., Vinyard, B.T., Baker, M.C., Tucker, C.S., 1995. Testing the temporal and spatial validity of site-specific models derived from airborne remote sensing of phytoplankton. Can. J. Fish. Aquat. Sci. 52, 1094–1107.

Milne, L.K., 1995. Feature selection with neural networks with contribution measures. In: Yao, X. (Ed.), Proceedings of the Eighth Australian Joint Conference on Artificial Intelligence (AI'95). World Scientific, Singapore, pp. 215–221.

Montana, D., 1995. Neural network weight selection using genetic algorithms. In: Goonatilake, S., Khebbal, S. (Eds.), Intelligent Hybrid Systems. Wiley, London, pp. 85–104.

Murray, A.G., Parslow, J.S., 1999. The analysis of alternative formulations in a simple model of a coastal ecosystem. Ecol. Model. 119, 146–166.

Odum, H.T., 1956. Primary production in flowing waters. Limnol. Oceaogr. 1, 102–117.

Olden, J.D., 2000. An artificial neural network approach for studying phytoplankton succession. Hydrobiologia 436, 131–143.

Olden, J.D., Jackson, D.A., 2002. Illuminating the ''black box'': understanding variable contributions in artificial neural networks. Ecol. Model. 154, 135–150.

Özesmi, S.L., Özesmi, U., 1999. An artificial neural network approach to spatial habitat modeling with interspecific interaction. Ecol. Model. 116, 15–31.

Paerl, H.W., 1988. Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. Limnol. Oceanogr. 33, 823–847.

Paerl, H.W., Pinckney, J.L., Fear, J.M., Peierls, B.L., 1998. Ecosystem responses to internal and watershed organic matter loading: consequences for hypoxia in the eutrophying Neuse River Estuary, North Carolina, USA. Mar. Ecol. Prog. Ser. 166, 17–25.

Paerl, H.W., Pinckney, J.L., Fear, J.M., Peierls, B.L., 1999. Fish kills and bottom-water hypoxia in the Neuse River and Estuary: reply to Burkholder et al. Mar. Ecol. Prog. Ser. 186, 307–309.

Paerl, H.W., Valdes, L.M., Pinckney, J.L., Piehler, M.F., Dyble, J., Moisander, P.H., 2003. Phytoplankton photopigments as indicators of estuarine and coastal eutrophication. BioScience 53, 953–964.

Paerl, H.W., Dyble, J., Pinckney, J.L., Valdes, L.M., Millie, D.F., Moisander, P.H., Morris, J.T., Bendis, B., Piehler, M.F., 2005. Using microalgal indicators to assess human and climactically-induced ecological change in estuaries. In: Bortone, S.A. (Ed.), Estuarine Indicators. CRC Press, Boca Raton, FL, pp. 145–174.

Paerl, H.W., Bales, J.D., Ausley, L.W., Buzzelli, C.P., Crowder, L.B., Eby, L.A., Fear, J.M., Go, M., Peierls, B.L., Richardson, T.L., Ramus, J.S., 2001. Ecosystem impacts of three sequential hurricanes (Dennis, Floyd and Irene) on the US's largest lagoonal estuary, Pamlico Sound, NC, USA. Proc. Natl. Acad. Sci. 98, 5655–5660.

Perry, D.A., 1995. Self-organizing systems across scales. Trends Ecol. Evol. 10, 241–244.

Pigg, R.J., Millie, D.F., Steidinger, K.A., Bendis, B.J., 2004. Relating cyanobacterial abundance to environmental parameters in the lower St Johns River Estuary. In: Steidinger, K.A., Landsberg, J.H., Tomas, C.R., Vargo, G.A. (Eds.), Harmful Algae. Florida Fish and Wildlife Conservation Commission. Florida

Institute of Oceanography, and Intergovernmental Oceanographic Commission of UNESCO, St. Petersburg, FL.

Pinckney, J., Dustan, P., 1990. Ebb-tidal fronts in Charleston Harbor, South Carolina: physical and biological characteristics. Estuaries 13, 1–7.

Pinckney, J.L., Millie, D.F., Howe, K.E., Paerl, H.P., Hurley, J.P., 1996. Flow scintillation counting of ^{14}C-labeled microalgal photopigments. J. Plankton Res. 18, 1867–1880.

Pinckney, J.L., Millie, D.F., Vinyard, B.T., Paerl, H.W., 1997. Environmental controls of phytoplankton bloom dynamics in the Neuse River Estuary (North Carolina, USA). Can. J. Fish. Aquat. Sci. 54, 2491–2501.

Pinckney, J.L., Paerl, H.W., Harrington, M.B., 1999. Responses of the phytoplankton community growth rate to nutrient pulses in variable estuarine environments. J. Phycol. 35, 1455–1463.

Principe, J.C., Euliano, N.R., Lefebrve, W.C., 2000. Neural and Adaptive Systems: Fundamentals Through Simulation. John Wiley and Sons Inc., New York, 656 pp.

Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1, 81–106.

Recknagel, F., French, M., Harkonen, P., Yabunaka, K.-I., 1997. Artificial neural network approach for modeling and prediction of algal blooms. Ecol. Model. 96, 11–28.

Richardson, A.J., Pfaff, M.C., Field, J.G., Silulwane, N.F., Shillington, F.A., 2002. Identifying characteristic chlorophyll *a* profiles in the coastal domain using an artificial neural network. J. Plankton Res. 24, 1289–1303.

Rudek, J., Paerl, H.W., Mallin, M.A., Bates, P.W., 1991. Seasonal and hydrological control of phytoplankton nutrient limitation in the lower Neuse River Estuary, NC. Mar. Ecol. Prog. Ser. 75, 133–142.

Sarnelle, O., 1992. Nutrient enrichment and grazer effects on phytoplankton in lakes. Ecology 74, 551–560.

Scardi, M., Harding Jr., L.W., 1999. Developing an empirical model of phytoplankton primary production: a neural network case study. Ecol. Model. 120, 213–223.

Schaffer, J.D., Whitley, D., Eshelman, L.J., 1992. Combinations of genetic algorithms and neural networks: a survey of the state of the art. In: Whitley, D., Schaffer, J.D. (Eds.), Proceedings of the International Workshop on Combinations of Genetic Algorithms and Neural Networks. IEEE Computer Society Press, Los Alamitos, CA, pp. 1–37.

Sigua, G.C., Steward, J.S., Tweedle, J.S., 2000. Water-quality monitoring and biological integrity assessment in the Indian River Lagoon, Florida: status, trends, and loadings (1988–1998). Environ. Manage. 25, 199–209.

Smith, A.S., Mason, A.K., 1997. Cost estimation predictive modeling: regression versus neural network. Eng. Econ. 42, 137–161.

Smith, D.W., Cooper, S.C., Sarnelle, O., 1988. Curvilinear density dependence and the design of field experiments on zooplankton composition. Ecology 69, 868–870.

Smith, S.V., Hollibaugh, J.T., 1993. Coastal metabolism and the oceanic organic carbon balance. Rev. Geophys. 31, 75–89.

Solorzano, L., 1969. Determination of ammonium in natural waters by the phenol-hypochlorite method. Limnol. Oceanogr. 14, 799–800.

Soyupak, S., Chen, D.-G., 2004. Fuzzy logic model to estimate seasonal pseudo steady state chlorophyll-*a* concentrations in reservoirs. Environ. Model. Assess. 9, 51–59.

Strickland, J., Parsons, T., 1972. A Practical Handbook of Seawater Analysis, Bulletin 167, 2nd ed. Fisheries Research Board Canada, Ottawa, Ont., Canada, 310 pp.

Swaney, D.P., Howarth, R.W., Butler, T.J., 1999. A novel approach for estimating ecosystem production and respiration in estuaries: application to the oligohaline and mesohaline Hudson River. Limnol. Oceanogr. 44, 1509–1521.

SYSTAT 10, 2000. Statistics 1. SPSS Inc, Chicago, 663 pp.

Tester, P.A., Varnum, S.M., Culver, M.E., Eslinger, D.L., Stumpf, R.P., Swift, R.N., Yungel, J.K., Black, M.D., Litaker, R.W., 2003. Airborne detection of ecosystem responses to an extreme event: phytoplankton displacement and abundance after hurricane induced flooding in the Pamlico-Albemarle Sound System, North Carolina. Estuaries 26, 1353–1364.

Thrun, S., 1995. Extracting rules from artificial neural networks with distributed representations. In: Tesauro, G., Touretzky, D., Leen, T. (Eds.), Advances in Neural Information Processing Systems 7. Massachusetts Institute of Technology Press, Cambridge, MA, pp. 505–512.

Walsh, J.J., Penta, B., Dieterle, D.A., Bissett, W.P., 2001. Predictive ecological modeling of harmful algal blooms. Hum. Ecol. Risk Asses. 7, 1369–1383.

Wetzel, R.G., 2001. Limnology: Lake and River Ecosystems, 3rd ed. Academic Press, San Diego, CA, 1006 pp.